UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

| APPLICATION NO. | FILING DATE | FIRST NAMED INVENTOR | ATTORNEY DOCKET NO. | CONFIRMATION NO. |
|---|---|---|---|---|
| 09/713,733 | 11/15/2000 | Mark Kantrowitz | 2942-991842 | 3519 |

7590 05/26/2004

Richard L Byrne
700 Koppers Building
436 Seventh Avenue
Pittsburgh, PA 15219-1818

| EXAMINER | |
|---|---|
| HUYNH, CONG LAC T | |

| ART UNIT | PAPER NUMBER |
|---|---|
| 2178 | 6 |

DATE MAILED: 05/26/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

PRG

| Office Action Summary | Application No. | Applicant(s) | |
|---|---|---|---|
| | 09/713,733 | KANTROWITZ, MARK | |
| | Examiner | Art Unit | |
| | Cong-Lac Huynh | 2178 | |

*-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --*

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

1)☒ Responsive to communication(s) filed on *15 November 2000*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

4)☒ Claim(s) *1-39* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-21,23, 26-30 and 37-39* is/are rejected.

7)☒ Claim(s) *22,24,25 and 31-36* is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

**Application Papers**

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

**Priority under 35 U.S.C. § 119**

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

**Attachment(s)**

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date *2, 3, 4*.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

## DETAILED ACTION

1. This action is responsive to communications: the application filed on 11/15/00, and the IDSs filed on 4/23/01, 6/3/02, 4/8/03 and 10/17/03.

2. Claims 1-39 are pending in the case. Claims 1, 37, 39 are independent claims.

### *Information Disclosure Statement*

3. The information disclosure statement filed 10/17/03 fails to comply with 37 CFR 1.98 since the copy of the IPER has not been considered as a prior art. It has been placed in the application file, but the information referred to therein has not been considered.

### *Claim Rejections - 35 USC § 112*

4. The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

5. Claims 15, 19, 23 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

Regarding claim 15, the claimed language is not proper when addressing that "the information retrieval method is a text classification method." Retrieving information is getting data from a database based on a request whereas text classifying is for putting text in different groups based on different criteria such as topic, type, etc. The two methods, therefore, are different, and one can not be the other.

Regarding claim 19, it is unclear what the document contains and how the two words relate to the user-specified minimum number of times when the claim states that "wherein the sequences of at least two words are considered as appearing in a document when *the document contains the sequence of at least two words at least a user-specified minimum number of times.*"

6. Claim 23 recites the limitation "wherein *the monotonic function* is the number of words in the phrase" in lines 1-2. There is insufficient antecedent basis for this limitation in the claim since claim 21 on which claim 23 is dependent, does not mention "the monotonic function." It is suggested that Applicants change the dependency of claim 23 to be dependent on claim 22, which mentions the monotonic function.

***Double Patenting***

7. Claim 20 is objected to under 37 CFR 1.75 as being a substantial duplicate of claim 19 since the only difference between the two claims is that claim 19 claims "the number of times" (line 4) and claim 20 claims "frequency" (line 4), where the number of times and the frequency have the same meaning. When two claims in an application are duplicates or else are so close in content that they both cover the same thing, despite a slight difference in wording, it is proper after allowing one claim to object to the other as being a substantial duplicate of the allowed claim. See MPEP § 706.03(k).

***Claim Rejections - 35 USC § 103***

8. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

9. Claims 1-2, 10-21, 26-29, 37-39 are rejected under 35 U.S.C. 103(a) as being unpatentable over Gomes et al. (US Pat No. 6,615,209 B1, 9/2/03, filed 10/6/00, priority 2/22/00).

Regarding independent claim 1, Gomes discloses:

- initially, selecting distinctive features contained in the collection of documents (col 3, lines 33-43, col 7, lines 43-56: the query-relevant parts extracted from the documents are distinctive features of the documents since the query-relevant parts includes specific information common to the documents; though Gomes does not explicitly mention the collection of documents, the fact that extracting the query-relevant parts from a plurality of documents suggests that these documents are in a collection for extracting)
- for each pair of documents having at least one distinctive features in common, comparing the distinctive features of the documents to determine whether the document are duplicate or near-duplicate document (col 3, line 33 to col 4, line 10, col 2, lines 38-56, col 7, lines 43-56: comparing each two documents for similarity based on the query-relevant parts referred as "snippets" where the

documents found can be duplicate, or duplicate with slight change, which means near-duplicate)

Gomes does not explicitly disclose that for each document, identifying the distinctive features contained in the document.

However, it would have been obvious to one of ordinary skill in the art at the time of the invention was made to have modified Gomes to include identifying the distinctive features contained in each document since the fact that the query-relevant parts are extracted from the plurality of documents suggests that the query-relevant parts are identified in each document before being extracted.

Regarding claim 2, which is dependent on claim 1, Gomes discloses that the method is applied to removing duplicates in document collections (figure 9, #930, col 8, lines 37-60: the duplicate removal management process uses query-relevant information to extract query-relevant information form documents indicates that Gomes method is applied for removing duplicates in a plurality of documents which are document collections).

Regarding claim 10, which is dependent on claim 1, Gomes discloses that the method is applied to creating a document index for use with a query system to efficiently find documents in response to a query which contains a particular phrase or excerpt (col 6, lines 10-27: "...a crawling process gets content from various sources accessible and stores such content...an automated indexing/sorting process may access the stored

content and may generate a content index ...a query processing process accepts queries and returns query results based on the content index..."; the returned query results based on the content index suggest that the query contains at least some words of the content index).

Regarding claim 11, which is dependent on claim 10, Gomes discloses that the document index can be utilized even if the particular phrase or excerpt was not recorded correctly in the document or in the query (col 6, lines 10-27: the fact that the queries are accepted and the query result are returned based on the content index suggests the document index can be used no matter how a particular phrase is recorded in the query or document).

Regarding claim 12, which is dependent on claim 1, Gomes discloses that the distinctive features appear in a different order in each of the documents (col 13, lines 1-22: "..the word frequencies of the query-relevant part ...two files with the same words *in different orders* would appear to be identical").

Regarding claim 13, which is dependent on claim 1, Gomes discloses the distinctive features are distinctive text fragments from the document in the document collection (col 7, lines 50-56; col 10, lines 56-67: the query-relevant information or the segments surrounding keyword occurrences are text fragments from the documents that show distinctive features of the document).

Regarding claim 14, which is dependent on claim 13, Gomes discloses that the method is applied to information retrieval methods (col 5, line 66 to col 6, line 2; col 7, lines 28-40).

Regarding claim 15, which is dependent on claim 14, Gomes does not disclose that the information retrieval method is a text classification method. However, it would have been obvious to an ordinary skill in the art at the time of the invention was made to have modified Gomes to include the fact that the information retrieval method is a text classification method since the process of retrieving information from a database would be performed faster based on the text classifying of documents in the database where the documents are stored according to various topics or types.

Regarding claim 16, which is dependent on claim 14, Gomes discloses that the information retrieval method assumes word independence, and the distinctive text fragments are added to an index set (col 6, lines 10-27: "...a crawling process gets content from various sources accessible and stores such content...an automated indexing/sorting process may access the stored content and may *generate a content index* ...a query processing process accepts queries and returns query results based on the content index..."; the returned query results based on the content index suggest that the content index contains at least some words of the query that is a distinctive text fragment).

Regarding claim 17, which is dependent on claim 13, Gomes discloses that the distinctive text fragments are sequences of at least two words that appear in documents in the document collection (col 10, lines 56-67: the segments or query-relevant information show that the distinctive text fragments are sequences of at least two words). Gomes does not disclose that the distinctive text fragments appear in a limited number of documents in the document collection. However, it would have been obvious to one of ordinary skill in the art at the time of the invention was made to have modified Gomes to include the limited number of documents in the document collection that the distinctive text fragments appear since only a number of documents having the text fragments that includes the keywords in the query, not all of the documents. Therefore, the number of documents is limited.

Regarding claim 18, which is dependent on claim 14, Gomes discloses that if one distinctive text fragments is contained within another distinctive text fragment within the same document, only the longest distinctive text fragment is considered as a distinctive feature (col 10, lines 44-67: the fact that segments surrounding keyword occurrences or keyword-in-context summaries suggest that the segment which is considered as the longest distinctive text fragment since it includes the query-related information, which is the shorter distinctive text fragments).

Regarding claim 19, which is dependent on claim 17, Gomes discloses that the sequences of at least two words are considered as appearing in a document when the

document contains the sequence of at least two words at least a user-specified minimum number of times (col 12, lines 18-35: the fact that a segment may be added to the query-relevant part QR only if it contains *at least a predetermined number of occurrences* of any of the keywords where a segment is a portion of a document suggests that the document contains the sequence of keywords and a specified minimum number of times of the occurrences of the keywords where it was obvious that the *predetermined number* of occurrences *can be defined* by user).

Regarding claim 20, which is dependent on claim 17, Gomes discloses that the sequences of at least two words are considered as appearing in a document when the document contains the sequence of at least two words at least a user-specified minimum frequency (col 12, lines 18-35: the fact that a segment may be added to the query-relevant part QR only if it contains *at least a predetermined number of occurrences* of any of the keywords where a segment is a portion of a document suggests that the document contains the sequence of keywords and a specified minimum frequency of the occurrences of the keywords where it was obvious that the predetermined number of occurrences can be defined by user).

Regarding claim 21, which is dependent on claim 17, Gomes discloses:

- the highest scoring sequences that are found in at least two documents in the document collection are considered distinctive text fragments (col 12, lines 40-54: the fact that only a predetermined number of the highest ranking segments

would be added to the query-relevant part QR suggests the highest ranking segments added to the query-relevant part QR be considered as distinctive text fragments)

Gomes does not explicitly disclose calculating a distinctive score for each sequence of at least two words. However, it would have been obvious to one of ordinary skill in the art at the time of the invention was made to have modified Gomes to include calculating a distinctive score for each sequence of at least two words since once a segment including sequences of words related to the query *is ranked as a highest ranking segments*, the ranking process must be carried out based on the scores of a plurality of segments. In other words, calculating a score for each sequence of words must be performed for the segment ranking.

Regarding claim 26, which is dependent on claim 17, Gomes does not explicitly disclose that the limited number of documents is selected by a user.

Instead, Gomes discloses that since the amount of text extracted influences a subsequent similarity measure, the tunable parameters 933 and 935 *should be adjusted* in concert (figure 9 and col 10, lines 44-50). Gomes further explains that "in general, the less information extracted, the more similar the documents may be found to be (so the similarity threshold should be set higher, or stated oppositely, the more information extracted, the less similar the documents may be found to be (so the similarity threshold should be set lower)" (col 10, lines 51-55).

It would have been obvious to one of ordinary skill in the art at the time of the invention was made to have modified Gomes to include the user's selection for the limited number of documents since the fact a user *can adjust the extraction parameters* or the similarity measure parameters for a desired result suggests a possibility for users to *select the limited number* of documents for the adjustment.

Regarding claim 27, which is dependent on claim 17, Gomes does not explicitly disclose that the limited number is defined by a linear function of the number of documents in the document collection.
However, as mentioned in claim 26 above, Gomes discloses that a user can select the parameters in the program to adjust text extraction and the similarity measure (col 10, lines 44-55).
It would have been obvious to one of ordinary skill in the art at the time of the invention was made to have modified Gomes to incorporate a linear function of the number of documents based on the adjusted parameters and the number of documents in the document collection.

Regarding claim 28, which is dependent on claim 17, Gomes discloses that the distinctive text fragments include glue words (col 10, line 56 to col 11, line 11: though the keywords preferably do not include the "stop word" or glue word such as "the", "it", "and", "or", etc. for the search, the keywords are included in the snippets, which are the

segments surrounding the keywords; therefore, the segments surrounding the keywords, that are equivalent to the distinctive text fragments, still include glue words).

Regarding claim 29, which is dependent on claim 17, Gomes does not explicitly disclose that the glue words do not appear at either extreme of the distinctive text fragments. However, it would have been obvious to one of ordinary skill in the art at the time of the invention was made to have modified Gomes to include the feature that the glue words do not appear at either extreme of the distinctive text fragments for the following reason. Since the glue words do not convey much information or convey some type of Boolean operations (col 11, lines 1-11), there is no need to include the glue words at the either extreme of the distinctive text fragments.

Regarding independent claim 37 and its dependent claim 38, Gomes discloses:

- initially, selecting distinctive features contained in the collection of documents (col 3, lines 33-43, col 7, lines 43-56: the query-relevant parts extracted from the documents are distinctive features of the documents since the query-relevant parts includes specific information common to the documents; though Gomes does not explicitly mention the collection of documents, the fact that extracting the query-relevant parts from a plurality of documents suggests that these documents are in a collection for extracting)
- for each pair of documents having at least one distinctive features in common, comparing the distinctive features of the documents to determine whether the

document are duplicate or near-duplicate document (col 3, line 33 to col 4, line 10, col 2, lines 38-56, col 7, lines 43-56: comparing each two documents for similarity based on the query-relevant parts referred as "snippets" where the documents found can be duplicate, or duplicate with slight change, which means near-duplicate)

Gomes does not explicitly disclose that for each document, identifying the distinctive features contained in the document.

However, it would have been obvious to one of ordinary skill in the art at the time of the invention was made to have modified Gomes to include identifying the distinctive features contained in each document since the fact that the query-relevant parts are extracted from the plurality of documents suggests that the query-relevant parts are identified in each document before being extracted.

Gomes also does not disclose that the method is applied to a collection of text spans where the text spans are sentences. However, it would have been obvious to one of ordinary skill in the art at the time of the invention was made to have utilized identifying duplicates and near-duplicates documents in Gomes to apply to identifying duplicates and near-duplicates text spans where text spans are sentences *since it was obvious that a document comprises a plurality of sentences*. Accordingly, the two documents are identified duplicates if they have the duplicate sentences. Therefore, comparing the distinctive features of the documents should be based on comparing the distinctive features of the text spans, which are sentences included in a document. In other words,

Gomes inherently includes identifying duplicate and near-duplicate text spans, which are sentences.

Independent claim 39 is for an apparatus of method claim 1, and is rejected under the same rationale.

10. Claims 3-7, 30 are rejected under 35 U.S.C. 103(a) as being unpatentable over Gomes as applied to claim 1 above, and further in view of Aiken (US Pat No. 6,240,409 B1, 5/29/01, filed 7/31/98).

Regarding claims 3 and 4, which are dependent on claim 1, Gomes does not disclose explicitly that the method is applied to detecting plagiarism and to detecting copyright infringement.

Aiken discloses a method for detecting the similarities between the two documents (abstract, col 3, lines 4-24) and applying the detecting of similarities for detecting plagiarism among a set of documents and providing copyright protection (col 18, lines 1-22).

It would have been obvious to one of ordinary skill in the art at the time of the invention was made to have combined Aiken into Gomes for the following reason. Aiken discloses applying the detecting of similarities of documents to detecting plagiarism and providing copyright protection, thus motivating to apply the duplicate determination in Gomes to detecting plagiarism and providing copyright protection since the duplicate

features of the two documents in Gomes are the same as the similarities between the documents in Aiken and copyright protection is for preventing of the copyright infringement.

Regarding claim 5, which is dependent on claim 1, Gomes does not disclose explicitly that the method is applied to determine the authorship of a document.
As mentioned in claims 3-4 above, Aiken discloses applying the detecting of similarities for detecting plagiarism among a set of documents and providing copyright protection (col 18, lines 1-22).
It would have been obvious to one of ordinary skill in the art at the time of the invention was made to have combined Aiken into Gomes for the following reason. Aiken teaches applying detecting the document similarities to detecting plagiarism and providing copyright protection, thus motivating to determine the authorship of a document, especially the duplicate documents in Gomes since both plagiarism and copyright protection are for confirming the real author of a document.

Regarding claim 6, which is dependent on claim 1, Gomes does not disclose that the method is applied to clustering successive versions of a document from among a collection of documents.
Aiken discloses clustering successive versions of a document from among a collection of documents (figures 1a-b, 4a and col 10, line 4 to col 11, line 46).

It would have been obvious to one of ordinary skill in the art at the time of the invention was made to have combined Aiken into Gomes since Aiken discloses clustering documents *based on similarities of the document contents* thus motivating to utilize the duplicate and near-duplicate features of documents in Gomes for clustering documents in a collection.

Regarding claim 7, which is dependent on claim 1, Gomes does not disclose that the method is applied to seeding a text classification or text clustering algorithm with sets of duplicate or near-duplicate.

Aiken discloses clustering documents using a text clustering algorithm based on the matching of the documents in a collection (figures 1a-b, 4a, col 7, lines 17-35).

It would have been obvious to one of ordinary skill in the art at the time of the invention was made to have combined Aiken into Gomes since Aiken discloses a text clustering algorithm applied on the matched documents in a document collection thus motivating to utilize the duplicate features of documents in Gomes as the matching features of the documents for applying the clustering algorithm.

Regarding claim 30, which is dependent on claim 1, Gomes does not disclose:

- counting the number of distinctive features in common
- wherein determining whether the pair of documents is duplicates or near-duplicates includes the steps of:

- for each pair of documents, calculating an overlap ratio by dividing the number of distinctive features in common by the smaller of the number of distinctive features per document
- comparing the overlap ratio to a threshold and if the overlap ratio is greater than the threshold, then the pair of documents are duplicates or near-duplicates, otherwise the pair of documents are not duplicates or near-duplicates

Aiken discloses a method for clustering documents based on detecting the similarities of the documents (abstract, figures 1a, 4a) where the similarities of each two documents are determined by:

- calculating an overlap ratio by dividing the number of distinctive features in common by the smaller of the number of distinctive features per document (col 11, lines 1-14: "The similarity of two documents is *defined by ratio C/T*, where *C is the number of hashes the two documents have in common* and *T is the total number of hashes taken of one of the documents*, which can be the current document or *the smaller document*...")
- counting the number of distinctive features in common (col 11, lines 1-14: calculating the ratio C/T inherently shows counting the number of distinctive features in common C)
- comparing the overlap ratio to a threshold and if the overlap ratio is greater than the threshold, then the pair of documents are duplicates or near-duplicates, otherwise the pair of documents are not duplicates or near-duplicates (col 11,

lines 15-46: "if C/T is less than the threshold (e.g. a predetermined parameter), the matches associated with the retrieved document are discarded ..." the fact that the matches are discarded if C/T is less than the threshold and only documents having an interesting or significant number of matches with the current document are retained suggests that the document having an significant number of matches with the current document have the *overlap ratio C/T greater than the threshold,* which means these documents are similar or duplicates to the current document, otherwise they are not)

It would have been obvious to one of ordinary skill in the art at the time of the invention was made to have combined Aiken into Gomes since Aiken teaches calculating the ratio C/T for determine the similarities or duplicates of documents providing the advantage of apply Aiken's calculating method for effectively determining the duplicates of the documents.

11. Claims 8-9 are rejected under 35 U.S.C. 103(a) as being unpatentable over Gomes as applied to claim 1 above, and further in view of Armstrong (US Pat No. 6,356,633 B1, 3/12/02, filed 8/19/99).

Regarding claims 8 and 9, which are dependent on claim 1, Gomes does not disclose that the method is applied to matching an email message with responses to the email message, and is to matching responses to an email message with the email message.

Armstrong discloses an email system that can access a database containing data and information related to predefined keyword lists, predefined response templates, predefined responses, etc., where the *keylists can be matched with the content of the fields associated with the email, such as the "TO", "FROM", "RE", date/time created, date/time sent, date/time received, and of course, the body of the email message itself* (col 5, lines 7-18, figure 2A, 3A-B).

It would have been obvious to one of ordinary skill in the art at the time of the invention was made to have combined Armstrong into Gomes since Armstrong discloses using the matching of keylists and the content of the email fields to detect the relationship between an email message and its response via the content of fields for sending and receiving message, thus motivating to utilize the document duplicate features of Gomes, where the duplicate features imply *matching of the documents based on a distinctive feature related to keywords in a query*, for matching an email and the response to the email and vice versa.

***Allowable Subject Matter***

12. Claims 22, 24-25, 31-36 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

13. Claim 23 would be allowable if rewritten to overcome the rejection(s) under 35 U.S.C. 112, second paragraph, set forth in this Office action and to include all of the limitations of the base claim and any intervening claims.

***Conclusion***

14. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Pugh et al. (US Pat No. 6,658,423 B1, 12/2/03, filed 1/24/01).

Turney (US Pat No. 6,470,307 B1, 10/22/02, filed 6/23/97).

Meyerzon et al. (US Pat No. 6,547,829 B1, 4/15/03, filed 6/30/99).

Razin et al. (US Pat No. 6,098,034, 8/1/00, filed 3/18/96).

Kathrow et al. (US Pat No. 6,263,348 B1, 7/17/01, filed 7/1/98).

Cullen et al. (US Pat No. 5,933,823, 8/3/99, filed 3/1/96).

Damerau et al. (US Pat No. 6,697,998 B1, 2/24/04, filed 6/12/00).

Chaney et al. (US Pat No. 6,104,990, 8/15/00, filed 9/28/98).

Cuomo et al. (US Pat No. 6,185,614 B1, 2/6/01, filed 5/26/98).

Kanza et al. (US Pat No. 5,258,910, 11/2/93, filed 9/2/92).

Hall (US Pat No. 6,643,686 B1, 11/4/03, filed 12/16/99).

Kirsch et al. (US Pat No. 6,070,158, 5/30/00, filed 11/13/97).

Ponte (US Pat No. 6,718,363 B1, 4/6/04, filed 7/30/99).

Broder et al. (US Pat No. 6,119,124, 9/12/00, filed 3/26/98).

Schuetze et al. (US Pat No. 6,598,054 B2, 7/22/03, filed 10/19/99).

Belanger et al. (US Pat No. 6,628,824 B1, 9/30/03, filed 3/11/99).

Sako et al. (US Pat No. 6,546,490 B2, 4/8/03, filed 9/15/99).

Katariya et al. (US Pat No. 6,549,897 B1, 4/15/03, filed 12/17/98).

Katariya et al. (US Pat No. 6,473,753 B1, 10/29/02, filed 12/18/98).

Scheussler et al. (US Pat No. 6,366,950 B1, 4/2/02, filed 4/2/99).

Subbaroyan et al. (US Pat No. 6,442,606 B1, 8/27/02, filed 8/12/99).

Floratos et al. (US Pat No. 6,092,065, 7/18/00, filed 2/13/98).

Lee et al., Duplicate detection for symbolically compressed documents, IEEE September 1999, pages 305-308.

Doermann et al., The detection of duplicates in document image databases, IEEE 1997, pages 314-318.

Lopresti, Models and algorithms for duplicate document detection, IEEE September 1999, pages 297-300.

15. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Cong-Lac Huynh whose telephone number is 703-305-0432. The examiner can normally be reached on Mon-Fri (8:30-6:00).

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Heather Herndon can be reached on 703-308-5186. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Conglaihuynh

Cong-Lac Huynh
Examiner
Art Unit 2178
5/4/04